

二、申报项目

名 称	元学习驱动的自动驾驶跨域场景视觉感知方法研究		
研究类别	基础研究 <input checked="" type="checkbox"/> 应用基础 <input type="checkbox"/> 技术开发 <input type="checkbox"/>		
项目来源	中国博士后科学基金面上资助项目 <input checked="" type="checkbox"/> 国家自然科学基金项目 <input type="checkbox"/> 国家社科基金项目 <input type="checkbox"/> 其他国家级项目 <input type="checkbox"/> 自选项目 <input type="checkbox"/>		
项目所属一级学科	计算机科学与技术	项目所属二级学科	计算机应用技术
交叉一级学科		交叉二级学科	
关键词	(限 5 个) 视觉感知，自动驾驶，连续变化域，离散变化域，元学习		
<div>1. 项目简介（限 500 字）</div> <p>背景：视觉感知是自动驾驶领域的关键技术之一，确保其通用性和可靠性至关重要。然而，复杂的天气环境变化极易影响视觉感知模型的推理能力，导致模型无法持续提供可靠的决策信息，这给自动驾驶的安全应用带来了极大的威胁。</p> <p>问题：复杂环境中采集的视觉数据通常存在明显的分布差异，这些差异可以根据变化规律划分为<u>离散变化域</u>（例如白天和黑夜）和<u>连续变化域</u>（例如亮度逐渐变化的环境）。现有的视觉感知模型难以同时降低不同变化类型的域差异，影响了模型抽取特征的判别力和通用性，从而降低了跨域场景的视觉感知精度。</p> <p>方法：本项目针对问题中的两种域差异进行深入研究：（1）针对连续域，将分布拉伸和压缩步骤嵌套到元学习的内外循环中，通过交替优化策略聚合连续域的分布中心；（2）针对离散域，利用生成的潜在域数据促进离散空间的知识传递，并通过多配对元学习框架降低离散域间的迁移难度；（3）针对复合域，利用元适配器建模域差异的变化规律，在适应新域时保留历史学习的跨域知识，以提高模型适应连续和离散域的广义泛化能力。</p> <p>展望：最终构建出一个可靠的高精度跨域视觉感知体系，为视觉感知技术在自动驾驶领域的产业化落地提供有力保障，并为深化可信人工智能的理论研究提供新的思路。</p>			
<div>2. 研究内容（限 2000 字。研究项目主要内容、主要创新点、项目的预期目标、科学意义、应用前景等。）</div> <div>2.1 研究背景和意义</div> <p>人工智能的快速进步给智能交通带来了新的发展契机，同时也掀起了全球范围内对自动驾驶系统的新一轮研发热潮。在自动驾驶领域，视觉感知是人工智能应用的关键技术之一。纯视觉技术凭借其高性能和低成本等优势，是目前多数自动驾驶厂商（如特斯拉）研</p>			

发感知系统的主流技术路线。如图2.1所示，视觉感知技术从采集的视觉数据中挖掘出丰富的视觉线索，可以用于辅助自动驾驶系统实现对车辆的实时、连续控制。然而，复杂的环境变化极易影响模型的推理能力，使其无法持续地提供可靠的高精度视觉感知结果，从而给自动驾驶系统的安全性带来极大挑战。上述问题与视觉感知模型的跨域泛化能力密切相关，因此，如何通过提高模型的广义泛化能力，使其更好地适应自动驾驶的复杂环境变化，具有重要的研究意义和应用价值。

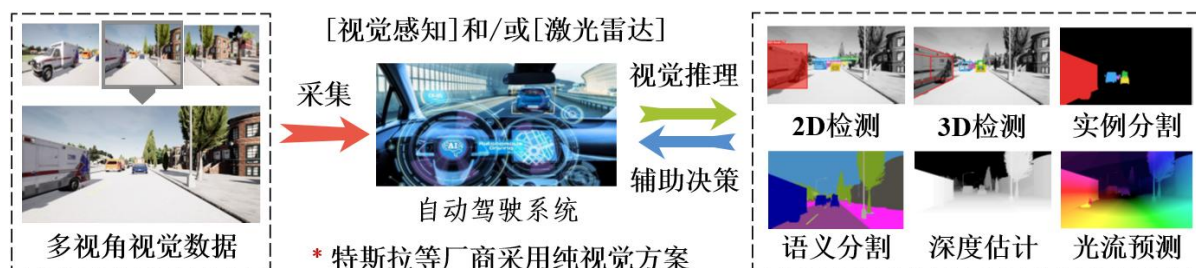


图2.1 基于视觉感知技术的自动驾驶系统

随着硬件效能的逐渐提高和数据资源的不断丰富，以深度学习为代表的人工智能技术推动了视觉感知的快速进步。然而，这种数据驱动的第二代人工智能在可解释性、鲁棒性等方面存在缺陷，导致安全事件频繁发生，这种问题在自动驾驶的视觉感知系统中也普遍存在。实际应用场景复杂环境变化导致采集视觉数据的分布存在明显差异，利用一种环境数据训练的深度学习模型通常无法直接适用于其他环境的数据。以图2.2为例，通过白天采集的数据训练模型适用于白天数据（C1可以区分红色三角形和绿色方块），但无法为夜间数据提供具有判别力的特征（C1无法区分蓝色圆形和黄色菱形），导致夜间环境的视觉感知精度严重下降。反之，夜间训练的模型也无法直接适用于白天采集的视觉数据。

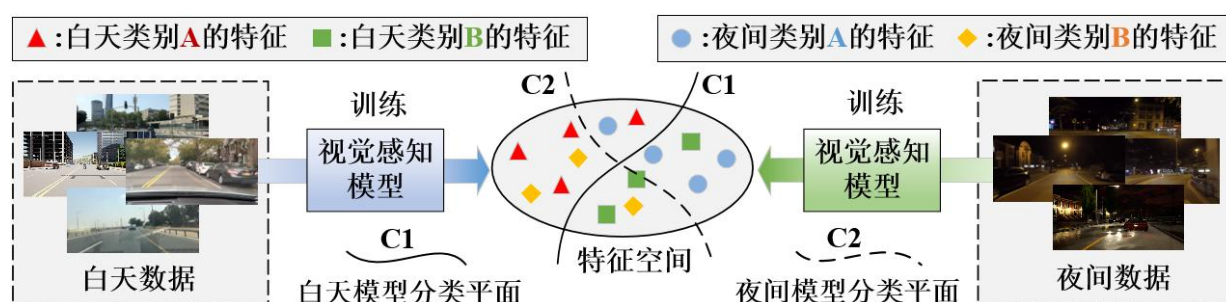


图2.2 跨域视觉感知示例

根据数据的分布差异，不同环境采集的数据可以被归纳到不同的域。考虑到域间差异的变化规律，可以将视觉数据划分为离散变化域和连续变化域。与连续域相比，离散域间的域差异无规律性且变化更大。这两种变化域在自动驾驶场景中极其常见，均会影响视觉感知模型的推理能力。为了解决该问题，许多研究者已经致力于提高视觉感知技术的通用性和可靠性。然而，现有方法普遍存在着广义泛化能力偏弱的问题，无法保障模型同时适应不同类型的跨域数据。因此，本项目将从剖析数据分布规律出发，利用元学习提高视觉感知模型的跨域泛化能力，以保障模型可以在复杂环境中持续地提供可靠的推理结果。

2.2 国内外相关研究现状

自动驾驶系统对安全性有着严格的要求，这就意味着视觉感知模型需要具有出色的泛化能力，能够在复杂环境中持续地提供可靠的视觉感知推理结果。域泛化一直以来是深度学习领域的研究热点，而跨域视觉感知技术凭借其在自动驾驶领域极高的应用价值，也引起了大批研究者的广泛关注。如图2.3所示，本项目将从两个方面介绍相关的研究工作。

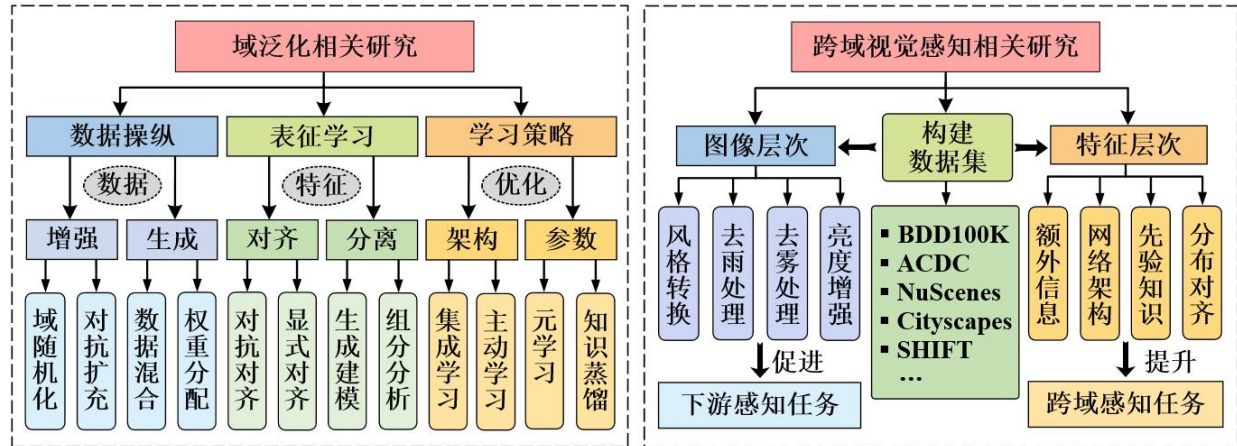


图2.3 国内外相关研究的主流技术路线

（1）域泛化的相关研究

数据操纵包括数据增强和生成方法，旨在促进模型从更多样化的数据中挖掘丰富的跨域知识。在有限的训练数据条件下，域随机化^[1]通过随机策略生成新的数据，从而模拟出更加复杂的训练环境。在保障数据可靠性的前提下，对抗扩充^[2]采用生成对抗方式扩充数据规模，以提高模型对域变化的鲁棒性。数据生成方法^[3]则通过自编码器等模型将多样的训练样本和对应的伪标签融合到一起，从而提高模型的通用性。为了有效利用额外的生成数据，一些方法^[4, 5]通过正则化等手段自适应地为训练样本分配权重。

表征学习主要包括特征对齐和分离方法。特征对齐学习旨在降低从不同域挖掘特征的分布差异。根据对齐策略，可以大致分类为对抗方法和显式方法。对抗对齐方法^[6, 7]利用生成器挖掘可以欺骗判别器的特征。通过交替优化生成器和判别器，促进模型挖掘出对域变化不敏感的特征表示。显式对齐方法主要通过降低最大均值差异^[8]或最优传输距离^[9]实现跨域特征对齐。分离表征学习则考虑从特征的差异化角度进行研究，主要包括生成建模和组分分析两类方法。生成建模^[10, 11]从样本和标签层面实现了跨域特征的解耦合，组分分析^[12]通过发挥域共享和域特定特征的协同优势来提高模型的泛化能力。

学习策略在提高模型泛化能力方面扮演着重要角色，主流的学习方法从网络架构或者模型参数优化的角度入手进行研究。集成学习^[13]将群体决策应用到级联的多个网络上，保证源域性能的基础上提高模型在不同目标域数据上的性能。元学习^[14]是一种指导模型学习如何学习的优化方法，与提升模型泛化能力的目标紧密相关。一些方法^[15, 16]通过元学习的双层交替优化策略，驱动模型从源域更好地迁移到目标域。此外，知识蒸馏^[17]以及主动学习^[18]等优化策略也经常用于提升模型的泛化能力。

（2）跨域视觉感知的相关研究

在自动驾驶系统的实际应用场景中，复杂环境变化引起的域差异问题会影响视觉感知模型的性能。为保证自动驾驶的安全性，视觉感知模型必须能够持续稳定地提供高精度的推理结果。为此，一些研究人员^[19-22]构建了包含复杂天气变化的自动驾驶数据集，为探索视觉感知技术的跨域泛化能力奠定了基础。对于复杂环境中采集的质量退化图像，一种简单的策略是在图像层次上降低域间差异，例如，通过迁移模型转换图像风格^[23]，或者通过去雨^[24]，去雾^[25]，亮度增强^[26]等底层视觉技术可以改善图像的质量。然而，这些方法只适用特定环境中采集的图像，并不适用于更加多样化的退化图像，特别是在复杂环境中。因此，一些方法^[27, 28]提出了普适性的增强方法，可以用于提高不同恶劣天气图像的视觉质量。但是，这些基于图像层面的预处理方法会增加额外的计算成本，进而影响整个视觉感知模型的推理速度。为了解决该问题，一些方法从特征层次入手，以端到端的学习方式提高模型的跨域能力。Ahmed等人^[29]通过先验知识提高恶劣天气条件中的交通标志检测精度。Hassaballah等人^[30]构建了一个适用于复杂天气车辆检测的深层网络架构。Zhou等人^[31]提出了一种基于牛顿极值搜索的自适应阈值分配方法，可以在不同光照条件下精准地检测车道线。Huang等人^[32]引入场景语义特征增强检测器抵抗恶劣天气变化的鲁棒性。Xue等人^[33]提出了一种新颖的多特征融合方法，以提高低亮度环境中行人和车辆的检测精度。

与目标检测任务相比，密集预测任务^[34]提供了更加细粒度的视觉线索，这有助于辅助自动驾驶系统更好地感知周围环境变化。Lee等人^[35]通过交替优化雾通滤波器和语义分割模型，逐渐地降低不同雾浓度环境间的特征分布差异。Choi等人^[36]在保持域不变特征的基础上，消除造成域间分布差异的图像风格信息。Reddy等人^[37]将边缘一致性作为先验知识，促进模型适应不同环境采集的退化图像。相比于其他的环境变化，自动驾驶应用场景中的光照变化更加普遍。每天都会经历白天与黑夜的交替，采集图像的亮度也随着光照条件的不断变化而发生改变。Sakaridis等人^[38]提出一种基于课程学习的域适应框架，将语义分割模型迁移到亮度逐渐降低的夜间图像。Lengyel等人^[39]提出了一种基于零次学习的昼夜域适应方法，促进模型挖掘出对光照变化不敏感的鲁棒特征。Gao等人^[40]提出了一种基于知识蒸馏的跨域方法，提高了模型对目标域夜间图像的分割精度。Deng等人^[41]通过一种双层网络逐步降低不同亮度环境间的域差异，从全局图像和局部区域两种粒度上提高模型的推理能力。Wu等人^[42]提出一种单阶段的无监督域适应方法，利用光照重置网络对昼夜图像的亮度进行对齐，然后通过概率重加权策略提高模型对夜间图像的分割精度。

2.3 现有工作存在的问题

近年来，提高深度学习模型的泛化能力已经成为众多研究者关注的焦点问题。随着数据集的不断丰富和评价体系的逐渐完善，许多具有跨域泛化能力的视觉感知模型被相继提出。然而，这些模型距离满足安全性和通用性的应用需求仍有很大差距。主要原因是现有视觉感知模型的广义泛化能力有限，容易导致以下问题：（1）现有方法虽然可以将模型从源域迁移到目标域数据，从而提高目标域数据上的感知精度，但不可避免地会影响模型

的整体推理能力，甚至可能降低源域数据的感知精度；（2）现有方法在同时进行多域泛化时，极易导致模型学习的跨域知识发生偏移，使得模型过渡拟合到某些特定域的数据，从而无法很好地适应其他目标域数据；（3）现有方法未考虑到自动驾驶数据独特的分布特点，离散域和连续域差异对模型的感知能力造成不同程度的影响。因此，如何有效地降低不同种类的域差异，是一个重要但未得到充分探索的研究方向。

为了解决现有方法中存在的问题，本项目将通过分析自动驾驶跨域数据的分布特点，剖析不同类型域差异的变化规律。在此基础上，利用元知识和元表征学习技术，提高模型在离散域和连续域数据上的广义泛化能力。最终，本项目将构建出一种适用于复杂环境变化的跨域视觉感知系统，以确保模型持续地实现高精度视觉感知推理。因此，本项目具有极高的研究价值和广泛的应用前景，对推动视觉感知技术的产业化落地具有重要作用。

参考文献

- [1] J. Huang, D. Guan, A. Xiao, and S. Lu, "FSDR: Frequency space domain randomization for domain generalization", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp: 6891-6902, virtual, Jun. 2021.
- [2] H. Wang, C. Xiao, J. Kossai, Z. Yu, A. Anandkumar, and Z. Wang, "AugMax: Adversarial composition of random augmentations for robust training", in *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, pp: 237-250, virtual, Dec. 2021.
- [3] N. Dvornik, J. Mairal, and C. Schmid, "On the importance of visual context for data augmentation in scene understanding", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 43(6): 2014-2028, 2021.
- [4] 李翔, 陈硕, 杨健, "泛化界正则项: 理解权重衰减正则形式的统一视角", *计算机学报*, 44(10): 2122-2134, 2021.
- [5] C. Tian, H. Li, X. Xie, Y. Liu, and S. Wang, "Neuron coverage-guided domain generalization", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 45(1): 1302-1311, 2023.
- [6] W. Zhang, D. Xu, W. Ouyang, and W. Li, "Self-paced collaborative and adversarial network for unsupervised domain adaptation", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 43(6): 2047-2061, 2021.
- [7] M. Ebrahimi, Y. Chai, H. Zhang, and H. Chen, "Heterogeneous domain adaptation with adversarial neural representation learning: Experiments on E-commerce and cybersecurity", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 45(2): 1862-1875, 2023.
- [8] 何相腾, 彭宇新, "跨域和跨模态适应学习的无监督细粒度视频分类", *软件学报*, 32(11): 3482-3495, 2021.
- [9] C. Chuang, Y. Mroueh, K. Greenewald, A. Torralba, and S. Jegelka, "Measuring generalization with optimal transport", in *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, pp: 8294-8306, virtual, Dec. 2021.
- [10] H. Li, S. Wang, R. Wan, and A. Kot, "GMFAD: Towards generalized visual recognition via multilayer feature alignment and disentanglement", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 44(3): 1289-1303, 2022.
- [11] X. Xing, R. Gao, T. Han, S. Zhu, and Y. Wu, "Instance-invariant domain adaptive object detection via progressive disentanglement", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 44(8): 4178-4193, 2022.

- [12] Q. Liu, and Z. Wang, “Collaborating domain-shared and target-specific feature clustering for cross-domain 3D action recognition”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp: 137-155, Tel Aviv, Israel, Oct. 2022.
- [13] J. Wang, Y. Qian, F. Li, J. Liang, and Q. Zhang, “Generalization performance of pure accuracy and its application in selective ensemble learning”, *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 45(2): 1798-1816, 2023.
- [14] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, “Meta-learning in neural networks: A survey”, *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 44(9): 5149-5169, 2022.
- [15] M. Choi, J. Choi, S. Baik, T. Kim, and K. Lee, “Test-time adaptation for video frame interpolation via meta-learning”, *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 44(12): 9615-9628, 2022.
- [16] C. Wan, X. Shen, Y. Zhang, Z. Yin, X. Tian, F. Gao, J. Huang, and X. Hua, “Meta convolutional neural networks for single domain adaptation”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp: 4672-4681, LA, USA, Jun. 2022.
- [17] 邵仁荣, 刘宇昂, 张伟, 王骏, “深度学习中知识蒸馏研究综述”, *计算机学报*, 45(08): 1638-1673, 2022.
- [18] 刘贵松, 郑余, 解修蕊, 黄鹂, 丁浩伦, “一种基于损失预测的双主动域适应算法研究”, *计算机学报*, 48(03): 579-593, 2023.
- [19] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, T. Darrell, “BDD100K: A diverse driving dataset for heterogeneous multitask learning”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp: 2633-2642, WA, USA, Jun. 2020.
- [20] C. Sakaridis, D. Dai, and L. Gool, “ACDC: The adverse conditions datasets with correspondences for semantic driving scene understanding”, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp: 10745-10755, Montreal, Canada, Oct. 2021.
- [21] C. Diaz-Ruiz, Y. Xia, Y. You, J. Nino, J. Chen, J. Monica, X. Chen, K. Luo, Y. Wang, M. Emond, W. Chao, B. Hariharan, K. Weinberger, and M. Campbell, “Ithaca365: Dataset and driving perception under repeated and challenging weather conditions”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21351-21360, LA, USA, Jun. 2022.
- [22] T. Sun, M. Segu, J. Postels, Y. Wang, L. Gool, B. Schiele, F. Tombari, F. Yu, “SHIFT: A synthetic driving dataset for continuous multi-task domain adaptation”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21339-21350, LA, USA, Jun. 2022.
- [23] J. Lin, Z. Chen, Y. Xia, S. Liu, T. Qin, and J. Luo, “Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation”, *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 43(4): 1254-1266, 2021.
- [24] 张学锋, 李金晶, “基于双注意力残差循环单幅图像去雨集成网络”, *软件学报*, 32(10): 3283-3292, 2021.
- [25] C. Zhou, M. Teng, Y. Han, C. Xu, and B. Shi, “Learning to dehaze with polarization”, in *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, pp: 11487-11500, virtual, Dec. 2021.
- [26] Z. Zhu, Z. Li, R. Zhang, C. Guo, and M. Cheng, “Designing an illumination-aware network for deep image relighting”, *IEEE Trans. Image Process. (TIP)*, 31: 5396-5411, 2022.
- [27] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, “Image-adaptive YOLO for object detection in adverse weather conditions”, in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1792-1800, Virtual, Feb. 2022.

- [28] A. Kulkarni, and S. Murala, "WiperNet: A lightweight multi-weather restoration network for enhanced surveillance", *IEEE Trans. Intell. Transp. Syst. (TITS)*, 23(12), pp: 24488-24498, 2022.
- [29] S. Ahmed, U. Kamal, and M. K. Hasan, "DFR-TSD: A deep learning based framework for robust traffic sign detection under challenging weather conditions", *IEEE Trans. Intell. Transp. Syst. (TITS)*, 23(6), pp: 5150-5162, 2022.
- [30] M. Hassaballah, M. Kenk, K. Muhammad, and S. Minaee, "Vehicle detection and tracking in adverse weather using a deep learning framework", *IEEE Trans. Intell. Transp. Syst. (TITS)*, 22(7): 4230-4242, 2021.
- [31] Y. Zhou, Z. Wang, and J. Wang, "Illumination-resilient lane detection by threshold self-adjustment using Newton-based extremum seeking", *IEEE Trans. Intell. Transp. Syst. (TITS)*, 23(10): 18643-18654, 2022.
- [32] S. Huang, T. Le, and D. Jaw, "DASNet: Joint semantic learning for object detection in inclement weather conditions", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 43(8): 2623-2633, 2021.
- [33] T. Xue, Z. Zhang, W. Ma, Y. Li, A. Yang, and T. Ji, "Nighttime pedestrian and vehicle detection based on a fast saliency and multifeature fusion algorithm for infrared images", *IEEE Trans. Intell. Transp. Syst. (TITS)*, 23(9), pp: 16741-16751, 2022.
- [34] S. Vandenhende, S. Georgoulis, W. Gansbeke, M. Proesmans, D. Dai, and L. Gool, "Multi-task learning for dense prediction tasks: A survey", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 44(7): 3614-3633, 2022.
- [35] S. Lee, T. Son, and S. Kwak, "FIFO: Learning fog-invariant features for foggy scene segmentation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp: 18889-18899, LA, USA, Jun. 2022.
- [36] S. Choi, S. Jung, H. Yun, J. Kim, S. Kim, and J. Choo, "RobustNet: Improving domain generalization in urban-scene segmentation via instance selective whitening", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp: 11580-11590, virtual, Jun. 2021.
- [37] N. Reddy, A. Singhal, A. Kumar, M. Baktashmotlagh, and C. Arora, "Master of all: Simultaneous generalization of urban-scene segmentation to all adverse weather conditions", in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp: 51-69, Tel Aviv, Israel, Oct. 2022.
- [38] C. Sakaridis, D. Dai, and L. Gool, "Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 44(6), pp: 3139-3153, 2022.
- [39] A. Lengyel, S. Garg, M. Milford, and J. Gemert, "Zero-shot day-night domain adaptation with a physics prior", in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp: 4379-4389, Montreal, Canada, Oct. 2021.
- [40] H. Gao, J. Guo, G. Wang, and Q. Zhang, "Cross-domain correlation distillation for unsupervised domain adaptation in nighttime semantic segmentation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9903-9913, LA, USA, Jun. 2022.
- [41] X. Deng, P. Wang, X. Lian, and S. Newsam, "NightLab: A dual level architecture with hardness detection for segmentation at night", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp: 16917-16927, LA, USA, Jun. 2022.
- [42] X. Wu, Z. Wu, L. Ju, and S. Wang, "A one-stage domain adaptation network with image alignment for unsupervised nighttime semantic segmentation", *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 45(1), pp: 58-72, 2023.

2.4 主要研究内容

围绕自动驾驶场景的跨域视觉感知技术，本项目将以数据分析为出发点，分析复杂环境变化的数据分布特点，并考虑复合域中的域差异变化规律，通过分而治之的策略分别降低连续域和离散域差异。以元知识学习为扩展点，从属性变化较小的数据出发降低连续变化的域差异，以此方式促进模型更好地泛化到域差异变化更明显的离散域。以元表征学习为拓展点，集中解决复合变化域中不同类型的域差异，提升模型对连续域和离散域的适应能力。本项目的研究内容如图 2.4 所示，首先，针对连续域，将拉伸和压缩步骤分别嵌套到元学习的内外循环中，通过交替优化两个步骤聚合连续域分布中心；其次，针对离散域，利用中间域数据促进离散域间的知识传递，通过多配对的元学习框架降低离散域的迁移难度；最后，针对复合域，利用元适配器建模不同类型的域差异，在适应新域的同时保留历史学习的跨域知识，提高模型对复合域的泛化能力。

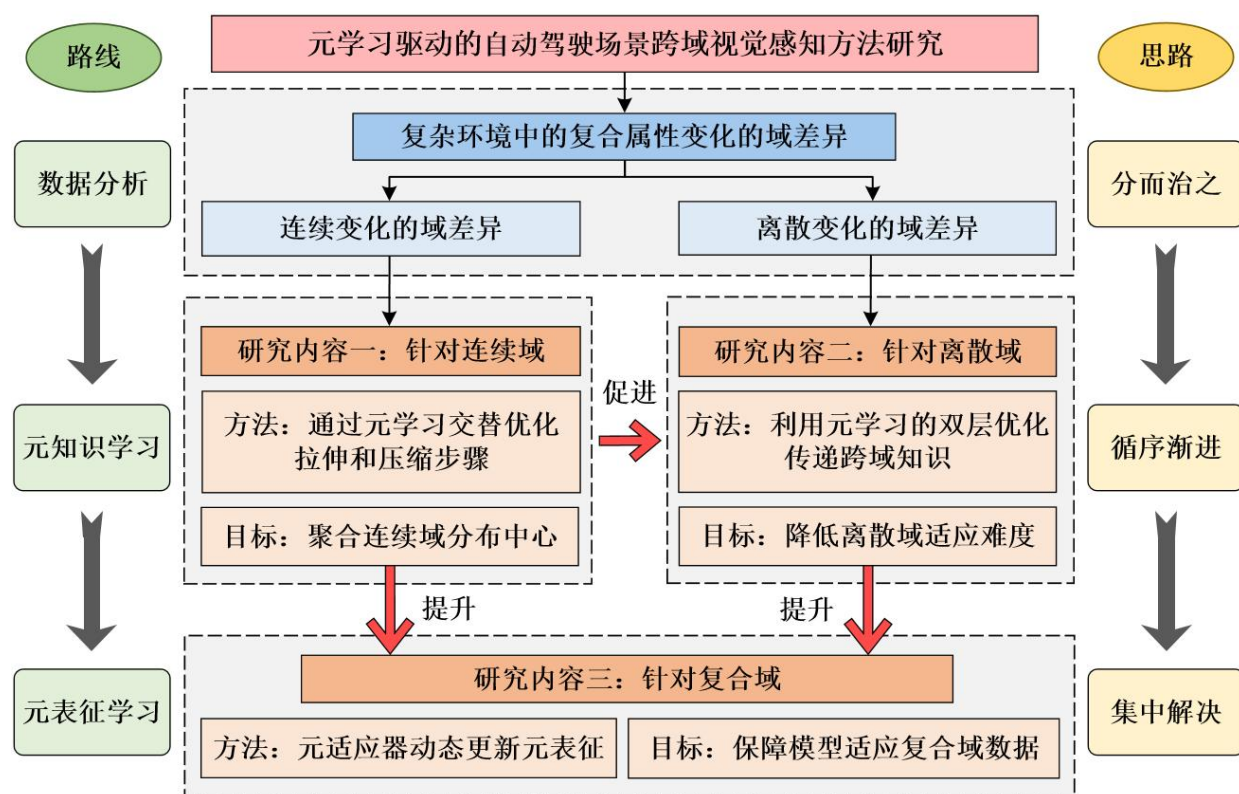


图 2.4 本项目的主要研究内容

在自动驾驶的实际应用中，不同环境采集的视觉数据存在明显的分布差异。相似环境中，服从相同分布的数据属于同一个域，而复杂环境通常包含多个不同的域。由于天气环境的变化具有规律性，自动驾驶采集的数据在不同域间的属性变化也呈现规律性。根据域差异变化的特点，可以将复杂环境包含的多个域归类为离散变化域和连续变化域。现有的自动驾驶场景跨域视觉感知模型不能同时有效地降低连续变化和离散变化的域差异，这意味着它们缺乏广义的泛化性。通过分析自动驾驶复杂环境中数据的分布特点，剖析不同类型域间属性的变化规律，将为提升视觉感知模型的跨域能力奠定良好的基础。本项目将以

数据分析为出发点，利用元知识和元表征学习构建一套高精度的跨域视觉感知体系，以保障模型可以在复杂环境中持续地提供可信赖的视觉感知结果，具体的研究内容包括：

（1）针对连续域，提出基于元学习拉伸和压缩循环优化的跨域视觉感知方法。该方法在高维特征空间中对数据分布进行拉伸和压缩，以聚合连续域分布中心，并降低目标域和源域间的域差异。在此基础上，将拉伸和压缩步骤分别嵌套到元学习的内外循环中，通过交替更新这两个步骤学习出跨域知识，从而指导模型更好地适应连续变化域数据。

（2）针对离散域，提出基于潜在域多配对元学习的跨域视觉感知方法。对于离散分布的源域和目标域数据，该方法利用风格迁移模型生成潜在域数据，并将其作为离散域间知识传递的桥梁，利用元训练从源域和潜在域之间学习元知识，在元测试阶段利用元知识指导模型更好地从潜在域迁移到目标域数据，从而降低模型在离散域之间迁移的难度。

（3）针对复合域，提出基于元表征动态更新的跨域视觉感知方法。该方法利用元适配器挖掘不同类型域间差异变化的特点，为模型提供跨域相关的元表征，从而更好地适应不同域数据。同时，该方法将关注跨域知识的遗忘问题，通过动态更新元表征，在适应新域的同时保留历史学习的跨域知识，同时地提高模型对连续域和离散域的泛化能力。

2.5 主要创新点

跨域视觉感知是一个重要且富有挑战性的研究方向，本项目以数据分析为出发点，揭示了视觉感知技术在复杂环境中性能下降的机理。并以此为基础，围绕可信人工智能的研发，本项目创新性地结合了元知识学习和元表征学习，提出了自动驾驶场景高精度的跨域视觉感知方法，为跨域视觉感知技术在自动驾驶领域的产业化落地奠定坚实的基础。与现有方法相比，本项目具有以下创新点：

（1）科学选题创新：自动驾驶复杂的环境变化给实现高精度视觉感知带来了极大的难度，同时，无法保障可信赖的感知推理结果也限制了视觉感知技术在自动驾驶领域的应用价值。本项目聚焦于自动驾驶系统实际的应用问题，创新性地考虑跨域数据自身独特的分布特点，能够有效地提高模型同时应对复杂环境中不同域的适应能力，从而保障模型在自动驾驶复杂环境中持续地提供可信赖的高精度视觉感知结果。

（2）研究思路创新：本项目以数据分析为出发点，剖析跨域数据的分布差异，探索域间属性的变化规律，为降低不同类型域差异奠定基础。以元知识学习为拓展点，在高维空间拉近连续域间的分布中心，并降低离散域间迁移模型的难度。以元表征学习为突破点，同时降低离散和连续变化在内的复合域差异。随着研究层次逐渐递进，研究内容不断深入，本项目最终将构建一种自动驾驶场景中可靠的高精度跨域视觉感知系统。

2.6 预期目标

（1）理论成果：本项目将构建国内自动驾驶的复杂环境数据集，以提高连续变化环境的数据体量，并完成离散化的跨域数据资源整合。然后，将提出基于元学习拉伸和压缩循环优化的跨域视觉感知方法，并逐步扩展到离散域适应，提出基于潜在域多配对元学习的

跨域视觉感知方法。并以此为基础，提出基于元表征动态更新的跨域视觉感知方法，以提升模型对复合域的泛化能力。最终，构建出一种适用于跨域场景的高精度视觉感知系统。

(2) 学术论文：围绕本项目的相关研究，将在本领域高水平期刊上发表 5-7 篇学术论文，其中至少包含 3 篇 IEEE Trans. 汇刊论文。

(3) 学术会议：参加 2-4 次国内外学术会议，与其他学者进行交流，并获取本领域最新的研究进展。

(4) 知识产权：申请国家发明专利 3 项以上，以推动本项目相关技术的产业化落地。

(5) 学生培养：协助培养 4-6 名具有一定科研素养的研究生，形成一个专注于自动驾驶场景跨域视觉感知技术的研究团队。

2.7 科学意义和应用前景

人工智能技术的快速发展给智能交通领域带来了新的发展机遇，但同时也带来了诸多挑战。全球范围内，自动驾驶技术的研究与开发正在蓬勃发展，这项技术的成功应用将为社会发展带来诸多好处，如提供更加顺畅的出行方式和物流解决方案。通过与人工智能、信息通讯等技术深度融合，自动驾驶技术正在重塑汽车及相关智能产业的生态体系。现代汽车产业正在向“传统工业与新一代信息产业的结合”转变，人工智能在自动驾驶领域的核心作用愈发明显，受到学术界和产业界的广泛关注。近年来，新一轮人工智能发展为计算机视觉技术带来了巨大的性能提升，计算机视觉技术开始在自动驾驶中广泛应用，这使得自动驾驶技术成为人工智能最早实现产业化落地的领域之一。然而，随着自动驾驶技术的发展和应用，也偶有交通事故发生，安全性成为了自动驾驶发展的首要考虑因素，也是保障人工智能技术产业化落地的关键因素之一。

视觉感知是人工智能在自动驾驶研究中的关键技术之一，在实际应用场景中，视觉感知技术为自动驾驶系统提供了丰富的视觉线索，可以辅助自动驾驶系统更好地感知周围环境变化，及时且准确地对车辆行驶做出控制。近年来，基于深度神经网络的深度学习方法逐渐兴起，加速了视觉感知技术在自动驾驶领域的快速发展。然而，基于深度学习的视觉感知方法在实际应用中却面临着诸多挑战。受到恶劣天气和复杂气候变化的影响，不同环境中采集的视觉数据可能存在域差异，现有的视觉感知模型难以在跨域数据上进行精准的预测，无法满足自动驾驶系统对安全性和可靠性的严格要求。究其根本原因在于，现有深度学习方法不具备对复杂环境变化的抵抗能力，跨域视觉感知模型普遍存在广义泛化能力偏弱的问题，不能有效地将模型同时迁移到不同类型的跨域数据，从而无法保障模型持续稳定地提供可信赖的高精度推理结果。为了解决上述问题，本项目将深入探索和研究自动驾驶场景跨域的视觉感知方法。通过分析跨域数据分布来挖掘属性变化规律，借助元知识学习和元表征学习相互协作的优势，同时解决不同类型跨域数据造成的分布差异问题，构建一套自动驾驶场景高精度的跨域视觉感知体系，这将为可信人工智能提供了一种新的研究思路，具有重要的研究价值和科学意义。此外，本项目的研究成果将为推动以视觉感知技术在自动驾驶领域的产业化落地打下坚实的基础，具有广阔的应用前景。

3. 研究方法（限 2000 字。研究计划、拟采取的研究方法、实验方案、技术路线、已具备的条件及目前进展的情况。）

3.1 技术路线

针对自动驾驶复杂环境下模型推理能力退化的问题，本项目将采用“前沿聚焦—模型设计—技术创新”的技术路线，构建出基于元学习的高精度跨域视觉感知方法，以提升模型的广义泛化能力，保障模型在不同环境中实现可信赖的视觉感知推理。具体来讲，本项目将以跨域数据分析为出发点，归纳域间数据的分布特点和域间属性的变化规律，为解决不同类型域差异问题奠定基础；以元知识学习为拓展点，通过双层的元优化策略拉近规律性变化的连续域差异，进一步渐进式地传递跨域知识，消除高度离散化数据间的分布差异；以元表征学习为突破点，针对包含离散和连续变化域差异的复合域，通过构建元适配器更新优化元表征，解决复合域泛化中存在的跨域知识遗忘问题，提高模型的广义泛化能力。

3.2 拟采取的方案

3.2.1 自动驾驶跨域数据的分析

在自动驾驶的实际应用场景中，不同环境采集的数据分布通常存在很大差异。将服从相同分布的数据定义为一个域，则复杂环境包含多个不同的域。域内数据服从相似的独立分布，域间的数据分布存在差异。如图 3.1 所示，自动驾驶场景的环境变化非常复杂，域间数据分布的属性变化存在不同种类，按照变化规律可以分为离散域和连续域。连续域的属性变化较小且更具有规律性，相比而言，离散域的属性变化通常无规律且更加剧烈。



图3.1 自动驾驶离散域和连续域示例

在深度学习方法中，模型通常在已知源域的数据上进行训练，然后将训练好的模型嵌入到自动驾驶系统，实现对未知目标域数据的视觉感知推理。当源域和目标域的数据分布存在差异，模型挖掘特征的判别能力会受到影响，从而降低模型的视觉感知精度。现有方法通过降低域差异将模型迁移到特定的目标域数据，但往往忽视了提升模型的广义泛化能力，即保障模型可以同时地适应不同环境采集的目标域数据。本项目将自动驾驶场景的跨域问题拆解为离散和连续的域差异问题，逐步降低域差异，从简单的连续变化数据扩展到高度离散化的，并通过解决知识的历史遗忘问题，提高视觉感知模型的广义泛化能力。

令 $P(x, y)$ 表示源域分布, $Q_t(x, y)$ 表示不断变化的目标域分布, 其中 t 为目标域轨迹上的采样位置。通过分布距离 $d(Q_{t_1}, Q_{t_2})$ 对变化属性进行量化, 则 $\lim_{\Delta t \rightarrow 0} d(Q_t, Q_{t+\Delta t}) = 0$ 表示属性的连续演化。域适应的目标是优化模型 f_θ 的参数 θ , 可以表示为如下形式:

$$\min_{\theta} \mathbb{E}_{t \sim U(0,1)} \mathbb{E}_{(x,y) \sim Q_t} L(f_\theta(x), y) = \min_{\theta} \int_0^1 \mathbb{E}_{(x,y) \sim Q_t} L(f_\theta(x), y) dt, \quad (1)$$

其中, L 表示损失函数。从 P 中采样 n_s 个样本, 构成源域训练集 $S = \{X_s, Y_s\} = \{x_s^i, y_s^i\}_{i=1}^{n_s}$, 其中 x_s^i 和 y_s^i 分别表示第 i 个图像和标签。目标域的测试集 $T = \{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ 可以看作随属性变化的分布轨迹, X_{t_i} 包含采样于 Q_{t_i} 的 n_{t_i} 个图像, $t_1 < t_2 < \dots < t_n$ 服从于均匀分布 $U(0,1)$ 。域间差异可以表示为: $d_{H\Delta H}(P, Q_t) = \sup_{\theta, \theta'} |\mathbb{E}_P L(f_\theta(x), f_{\theta'}(x)) - \mathbb{E}_{Q_t} L(f_\theta(x), f_{\theta'}(x))|$, 进一步可以衡量模型跨域学习的泛化能力: $\lambda_t = \min_{\theta} |\mathbb{E}_P L(f_\theta(x), y) - \mathbb{E}_{Q_t} L(f_\theta(x), y)|$ 。对于模型任意的参数 θ , 目标轨迹 t_1, t_2, \dots, t_n 的采样概率至少为 $1-\delta$, 则可知:

$$\mathbb{E}_t \mathbb{E}_{Q_t} L(f_\theta(x), y) \leq \mathbb{E}_P L(f_\theta(x), y) + \frac{1}{n} \sum_{i=1}^n [d_{H\Delta H}(P, Q_{t_i})] + \mathbb{E}_t \lambda + O\left(\frac{\alpha}{\delta n}\right), \quad (2)$$

其中, n 表示目标域轨迹的长度, α 可以看作是目标域属性的演化速率。为了泛化到属性不断演变的目标域, 需要考虑目标域轨迹上的随机样本, 而不能只关注如何将模型迁移到特定的单个目标域。自动驾驶场景的域差异问题具有独特的规律性, 实际应用场景中存在包括离散域和连续域在内的复合域分布状态, 且两种域差异在分布轨迹上的演化速率存在差别。如图 3.2 所示, 不同环境采集数据的分布轨迹存在差异。在训练阶段, 简单地混合不同类型的跨域数据极易造成跨域知识发生偏差, 给同时降低不同类型的域差异带来了极大的难度。由此可见, 在提高模型广义泛化力时需要区别对待离散域和连续域差异。

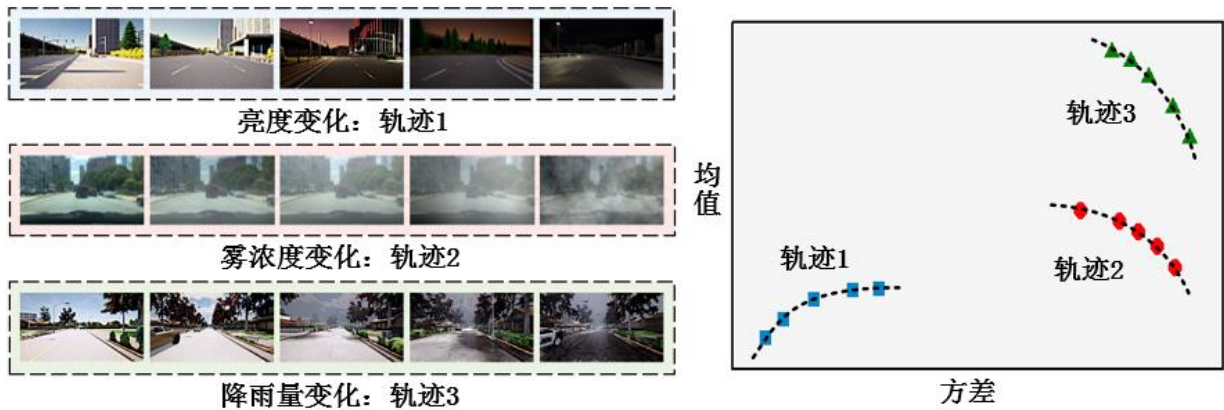


图3.2 不同域数据及对应轨迹的分布

现有的跨域视觉感知方法未能对自动驾驶场景的数据进行深入分析, 而直接将模型迁移到特定环境下的单个目标域。相比之下, 本项目通过深入剖析跨域数据, 将自动驾驶场景的复合域拆分为离散域和连续域, 首先从连续域适应入手, 通过元知识学习拉近域直之间的分布距离, 然后扩展到更加离散化的目标域数据, 通过潜在域降低模型跨离散域的元知识迁移难度, 最后通过元表征学习提高模型对不同分布类型目标域的广义泛化能力。

3.2.2 连续域：基于元学习拉伸和压缩循环优化的跨域方法

连续的变化域在自动驾驶场景中极其常见，例如，不同时刻的阳光光照不同，采集图像的亮度也随之变化，形成了关于亮度变化的连续变化分布轨迹。自动驾驶场景中还有很多类似的连续变化域，不同降雨量以及不同雾浓度等环境采集的视觉数据也存在相似连续变化的域差异。令 $P_{\theta(\alpha)}(x, y)$ 表示连续域的数据分布，图 3.3 展示了在高维空间上 $\theta(\alpha)$ 的分布情况。从 $P_{\theta(\alpha_i)}$ 中采样两个源域 $\{S_i(x, y)\}_{i=1}^2$ 以及 N 个目标域 $\{T_j(x)\}_{j=1}^N$ ，通过两个源域可以更好地建模未知目标域的变化，有助于保持模型在连续域数据上实现高精度的视觉感知。为此，本部分提出一种连续域适应框架以及一种包含拉伸和压缩交替优化的训练策略。

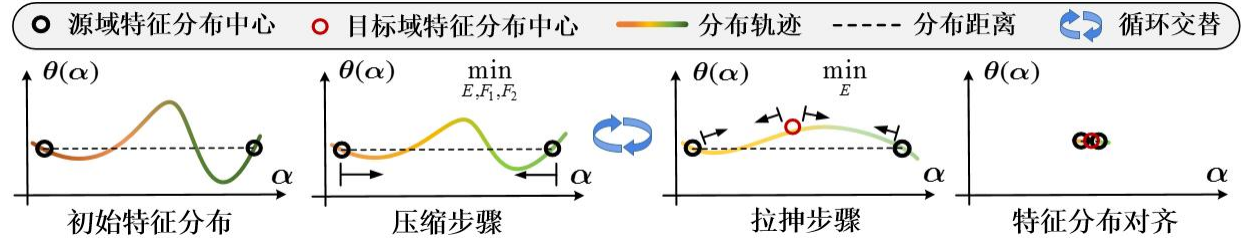


图 3.3 连续域适应的流程图

图 3.4 展现本部分搭建的连续域适应的框架。对于输入数据 x ，将利用特征编码器 E 提取特征 $e = E(x)$ 。连续变化域中通常存在若干个目标域，并且它们间的域差异不同，将通过源域的样本队列 Q_1 和 Q_2 挖掘出全局域差异，以此方式将域差异从小批量源域样本中解耦。此外，还将引入两个适应模块 A_1 和 A_2 分别作用于 Q_1 和 Q_2 。每个模块中， F_1 和 F_2 表示内容分类器， C_1 和 C_2 表示域判别器，特征 e_p 和 e_q 间的分布差异可以表示为如下形式：

$$D_j(e_p, e_q) = \sup_{C_j} -\mathbb{E}_{e_p \sim P} \log[F_j(e_p) \circ C_j(e_p)] - \mathbb{E}_{e_q \sim Q} \log[F_j(e_q) \circ C_j(e_q)], \quad (3)$$

其中， $j=1, 2$ ， \circ 表示组合函数。

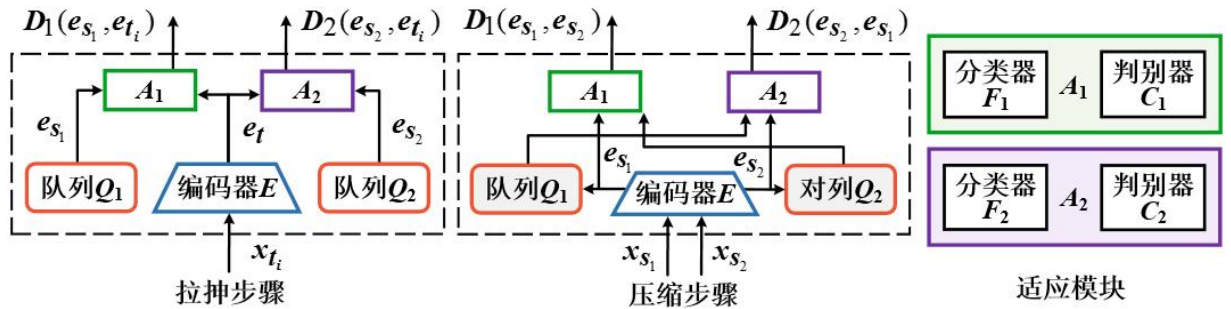


图 3.4 连续域适应的网络框架

拉伸步骤。从 S_1 ， S_2 和 T_i 中随机采样样本，通过优化 E 将目标域中心向源域分布轨迹上推动，该步骤可以通过最小化 $D_1(e_{s_1}, e_{t_i})$ 和 $D_2(e_{s_2}, e_{t_i})$ 的总和来实现，可以表示为：

$$\min_E \sum_{i=1}^N D_1(e_{s_1}, e_{t_i}) + D_2(e_{s_2}, e_{t_i}). \quad (4)$$

压缩步骤。从 S_1 和 S_2 中随机采样样本，通过优化 E ， F_1 以及 F_2 缩短源域轨迹长度，该步骤的优化目标可以表示为如下所示：

$$\min_{E, F_1, F_2} D_1(e_{s_1}, e_{s_2}) + D_2(e_{s_2}, e_{s_1}). \quad (5)$$

令 P_{s_j} 和 P_{t_i} 分别表示 e_{s_j} 和 e_{t_i} 的分布，可以推导出公式 (3) 中 $F(e) \circ C(e)$ 的最优值：

$$\text{拉伸步骤: } F_j(e) \circ C_j(e) = \frac{P_{s_j}}{P_{s_j} + P_{t_i}}, \quad (6)$$

$$\text{压缩步骤: } F_1(e) \circ C_1(e) = \frac{P_{s_1}}{P_{s_1} + P_{s_2}}, F_2(e) \circ C_2(e) = \frac{P_{s_2}}{P_{s_2} + P_{s_1}}. \quad (7)$$

交替优化的收敛证明：最小化公式(4)和(5)可以对齐源域和目标域，在公式(4)和(5)中分别替换公式(6)和(7)，优化目标可以表示为：

$$\min_E \begin{cases} \sum_{i=1}^N JS(P_{s_1} \parallel P_{t_i}) + JS(P_{s_2} \parallel P_{t_i}) & \text{拉伸步骤} \\ 2JS(P_{s_1} \parallel P_{s_2}) & \text{压缩步骤} \end{cases}, \quad (8)$$

其中，目标域和两个源域间的 JS 散度在拉伸步骤中被最小化，两个源域间的 JS 散度在压缩步骤中被最小化。在公式(8)中，拉伸步骤的解是 $p_{s_1}=p_{t_i}$ 和 $p_{s_2}=p_{t_i}$ ，压缩步骤的解是 $p_{s_1}=p_{s_2}$ 。两个步骤交替运行，当 $p_{s_1}=p_{s_2}=p_{t_i}$ 达到全局最小值，如图 3.4 中的分布轨迹最终将聚合为一个点。为了保障未知的目标域可以被拉近到源域的分布轨迹，在优化阶段加入了一种连续的约束。令探测目标域和其附近目标域的属性分别表示为 α 和 $\alpha+\Delta\alpha$ ，通过 E 提取的特征分别表示为 e_α 和 $e_{\alpha+\Delta\alpha}$ 。给定常量 η ， $D_i(e_{s_i}, e_{t(\alpha+\Delta\alpha)})$ 为如下形式：

$$D_i(e_{s_i}, e_{t(\alpha+\Delta\alpha)}) \leq \eta\Delta\alpha + D_i(e_{s_i}, e_{t(\alpha)}). \quad (9)$$

属性变化导致域分布变化，当 $\Delta\alpha \rightarrow 0$ 时， $e_t(\alpha + \Delta\alpha) = e_t(\alpha) + \beta\Delta\alpha$ ，其中 β 是另一个常量，公式(9)可以重新定义为：

$$\|\nabla_{e_t} D_i(e_{s_i}, e_t)\| \leq \eta / \|\beta\|. \quad (10)$$

设置 $\eta / \|\beta\| = 1$ 并将其作为梯度下降的惩罚约束，则公式(4)中拉伸步骤的优化目标可以转化成如下形式：

$$\min_E (\|\nabla_{e_t} D_i(e_{s_i}, e_t)\| - 1)^2. \quad (11)$$

通过最小化源域和探测目标域之间的差异 $D_i(e_{s_i}, e_t)$ ，可以隐式地降低源域和未知目标域之间的差异 $D_i(e_{s_i}, e_{t(\alpha+\Delta\alpha)})$ 。本部分提出的交替优化策略以一种随机的方式稳健地估计域差异，在元学习的优化过程中，将公式(11)和公式(5)对应的拉伸和压缩步骤分别嵌套进元训练的内循环和外循环，通过循环迭代优化出元知识。在元测试阶段，利用这些元知识可以指导模型更好地聚合连续变化数据中不同域的分布中心。

3.2.3 离散域：基于潜在域多配对元学习的跨域方法

自动驾驶场景的复杂环境中，通常存在多个不同的离散变化域。为了有效地降低离散域差异，本部分将提出一种基于潜在域的元学习方法。令包含 K 个源域和包含 N 个目标域的集合分别表示为 $D_S = \{D_S^0, D_S^1, \dots, D_S^{K-1}\}$ 和 $D_T = \{D_T^0, D_T^1, \dots, D_T^{N-1}\}$ 。在每次迭代中，从 D_S 和 D_T 中分别随机采样源域图像 X_S 和目标域图像 X_T 。然后，通过风格迁移网络生成潜在域图像 X_L ，用于降低 X_S 和 X_T 之间知识传递的难度。图3.5展示了离散域风格迁移网络以及生产的潜在域，可以看出，生成的 X_L 不但具有 X_S 的内容还具有 X_T 的风格。

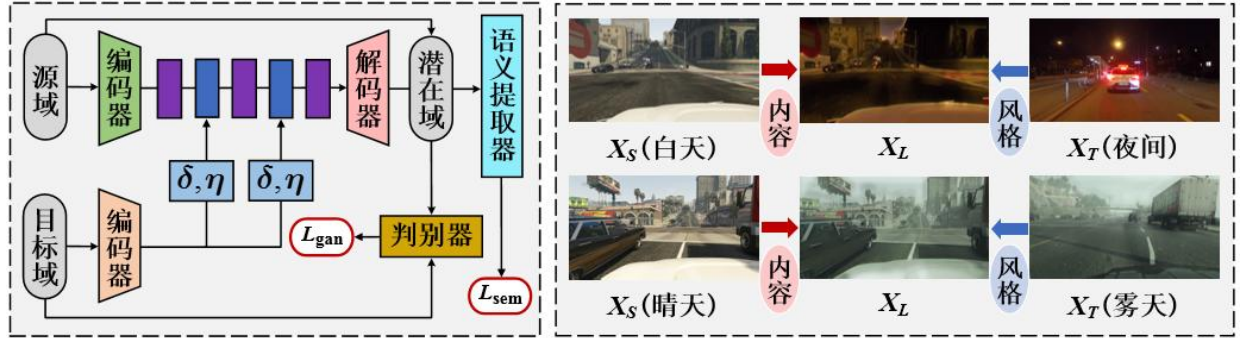


图3.5 离散风格迁移网络结构及潜在域图像示例

为了权衡迁移图像的风格和内容，离散风格迁移网络包含两个编码器，分别从源域和目标域提取风格和内容信息。在目标域对应编码器后面连接全连接层，学习出用于修正源域特征的仿射参数 (δ, η) ，可以表示为如下形式：

$$\hat{F}_z^S = \delta_z \left| \frac{F_z^S - \mu(F_z^S)}{\sigma(F_z^S)} \right| + \eta_z \quad (12)$$

其中， F_z^S 和 \hat{F}_z^S 分别表示第 z 层上修正前和修正后的源域特征图， σ 和 μ 分别表示对应特征图的均值和方差。令 f_p 表示经过预训练的语义分割模型，首先将 f_p 作为语义信息提取器，分别从源域和迁移后图像提取语义特征，然后通过语义一致性损失函数进一步约束迁移前后图像间语义内容的一致性，语义一致性损失函数可以表示为如下形式：

$$L_{sem}(X_S, X_L) = \sum_{h=1}^H \sum_{w=1}^W \left\| f_p(X_S)^{(h,w)} - f_p(X_L)^{(h,w)} \right\|, \quad (13)$$

其中， $H \times W$ 表示潜在域图像的尺寸。为了提升迁移后图像的视觉质量，将通过生成对抗损失优化迁移图像和目标域图像的相似度，可以表示为如下形式：

$$L_{gan}(X_T, X_L) = - \sum_{h=1}^H \sum_{w=1}^W y_d (1 - V(X_T^{(h,w)}))^2 + (1 - y_d) (1 - V(X_L^{(h,w)}))^2, \quad (14)$$

其中， V 表示风格判别器， $y_d \in 0,1$ 表示域标签，迁移的潜在域图像对应标签为1，目标域图像对应标签为0。对于从 D_S 和 D_T 中随机采样的源域 X_S 和目标域图像 X_T ，通过最小化损失函数 $L_{overall}$ 对风格迁移网络进行优化，可以得到高质量的潜在域图像 X_L ：

$$L_{overall}(X_S, X_T, X_L) = \lambda L_{sem}(X_S, X_L) + \beta L_{gan}(X_T, X_L), \quad (15)$$

其中， λ 和 β 分别表示损失函数的权重。对于离散分布的源域和目标域数据，生成的潜在域数据可以显著降低迁移模型的难度。图3.6展示本部分离散域适应的元学习过程，搭建的网络包含特征编码器 E 和域判别器 C ，二者之间通过梯度翻转层使 E 和 C 可以被端到端地训练。域适应损失函数 $L_{uda}=L_{task}+L_{adv}$ 可以作为单独配对域适应的优化目标，其中 L_{task} 和 L_{adv} 分别为感知损失和对抗损失。本部分提出的元学习优化包含两个域适应过程，在元训练阶段，首先通过元训练从源域 S 和任意两个潜在域 L_{k1} 和 L_{k2} 的配对中学习元知识，然后在元测试阶段，元知识将作为初始条件指导模型更好地泛化到目标域 T_{k1} 和 T_{k2} 。

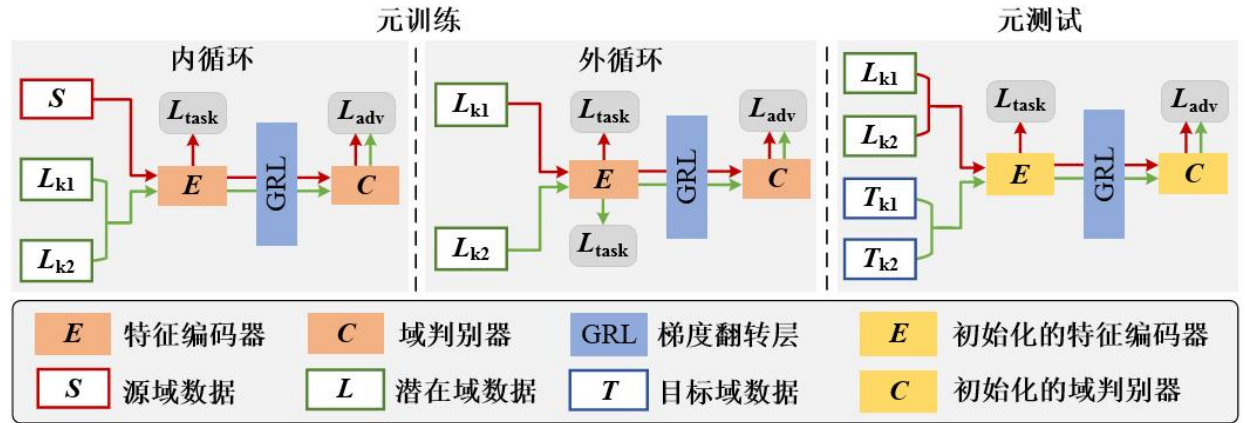


图3.6 离散域适应的元学习优化

在元训练阶段，可以将初始条件（即模型初始化参数）的元学习优化过程看作是一个包含内循环和外循环的双层优化问题，可以表示为如下形式：

$$\theta = \arg \min_{\theta} \overbrace{L_{out}(L_{in}(\theta; D_{support}); D_{query})}^{\text{外循环}}, \quad (16)$$

内循环

其中， $L_{inner}(\theta; D_{support})$ 和 $L_{out}(\theta^*; D_{query})$ 分别表示在支撑集 $D_{support}$ 和查询集 D_{query} 上计算的损失，外循环的初始参数 $\theta^* = \arg \min L_{inner}$ 是内循环更新后的模型参数。在内循环优化时，首先将 L_{inner} 替换成域适应损失 L_{uda} ，然后假设两个潜在域 $\overline{D_{L_{k1}}}$ 和 $\overline{D_{L_{k2}}}$ 无标签，使模型从源域 S 同时迁移到两个潜在域。在外循环优化时，将实际带有标签的潜在域作为 D_{query} ，通过 L_{task} 验证域适应性能，则双层循环的优化目标转化为学习初始的参数 θ_0 ，表示为如下形式：

$$\theta_0 = \arg \min_{\theta_0} \overbrace{L_{task}(L_{uda}(\theta_0; D_S, \overline{D_L}); D_L)}^{\text{外循环}}, \quad (17)$$

内循环

在元测试阶段，首先利用元训练学习的 θ_0 对编码器 E 和域判别器 C 进行初始化，然后进一步指导模型从潜在域 $D_{L_{k1}}$ 和 $D_{L_{k2}}$ 迁移到目标域 $D_{T_{k1}}$ 和 $D_{T_{k2}}$ 。本部分提出的多配对域适应框架可以降低离散域间知识传递的难度，将为进一步的复合域适应打下良好的基础。

3.2.4 复合域：基于元表征动态更新的跨域方法

针对连续域和离散域，本项目采用分而治之的策略，渐进式地降低域间分布差异。在此基础上，本部分将提出一种基于元表征动态更新方法，可以解决模型在不断适应新目标域的同时遗忘跨域知识的问题，并且有助于进一步提高模型对连续域和离散域的广义泛化能力。元表征的学习目标是挖掘出对跨域数据具有良好泛化性的特征表示，在元训练阶段，通过内循环最小化支持集上的误差，并且利用外循环最小化查询集上的误差，在元测试阶段，通过元适配器为新的目标域特征分配权重，促进模型适应不断变化的目标域数据。

如图 3.7 所示，本部分将搭建一个复合域适应框架，其中 E_θ 表示特征编码器， g_ϕ 表示域适配器， c_w 表示视觉感知模块。在不断变化的复合域中，随机采样两个目标域轨迹，分别作为元表征学习的支撑集 $T_{spt}=(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ 和查询集 $T_{qry}=(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ ，并且随机采样源域数据，并将其表示为 $S=\{X_s, Y_s\}=\{x_{s,i}, y_{s,i}\}_{i=1}^{n_s}$ 。

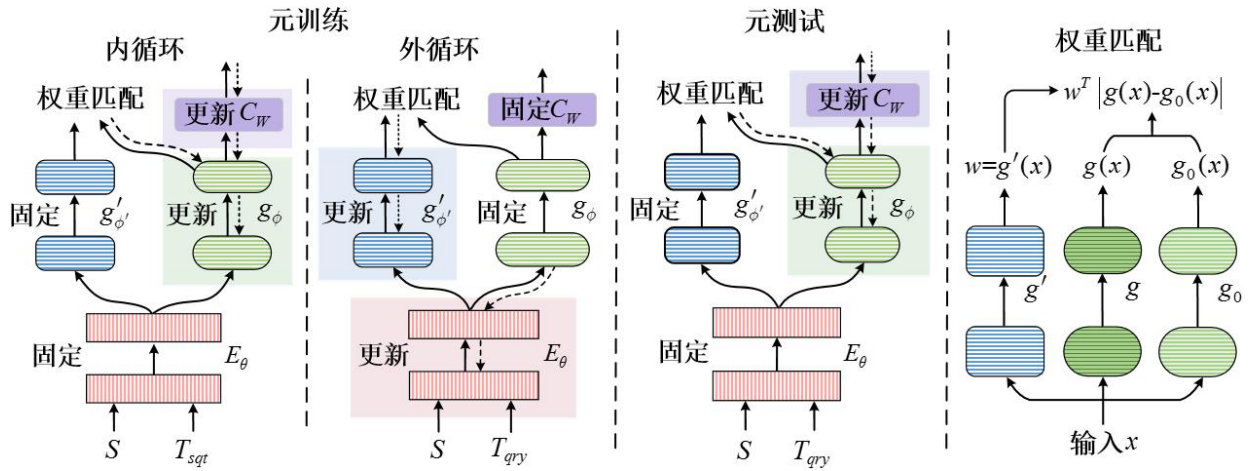


图 3.7 复合域的元表征学习框架

在内循环，通过训练优化适配器 g_ϕ 和视觉感知预测模块 c_w 的参数 ϕ 和 w ，其目标是促进从 E_θ 挖掘的特征可以适应不断演化的目标域数据。为此， E_θ 的参数在内循环是固定不变的，通过参数 ϕ_0 和 w_0 分别对 g 和 c 进行初始化之后，然后利用随机梯度下降方法将模型适应到序列 T_{spt} 中的目标域样本，参数更新的公式如下所示：

$$(\phi, w)_{i+1} \leftarrow (\phi, w)_i - \eta_{in} \nabla_{(\phi, w)} [L(f_{\theta, \phi_i, w_i}(X_s), Y_s) + d(f_{\theta, \phi_i, w_i}(X_s), f_{\theta, \phi_i, w_i}(X_{t_i}))], \quad (18)$$

其中， $i=0, 1, \dots, n-1$ ， η_{in} 表示内循环的学习率， L 表示视觉感知任务的损失函数， d 表示联合的最大均值距离，用于计算特征间的分布差异。

在外循环优化时，通过元表征学习保障内循环域适应的有效性，一个直接的策略是在查询集上最小化 3.1 部分公式(1)的损失。为了专门为复合域学习特有的元表征，需要把域属性变化速率 α 考虑在内，可以通过 $\max_i d(f_{\theta, \phi_n, w_n}(X_{t_{i-1}}), f_{\theta, \phi_n, w_n}(X_{t_i}))$ 近似 α ，然后通过公式 (19) 更新编码器 E_θ 的参数 θ ：

$$\begin{aligned} \theta \leftarrow \theta - \eta_{out} \nabla_{\theta} [L(f_{\theta, \phi_n, W_n}(X_s), Y_s) + \frac{1}{n} \sum_{i=1}^n d(f_{\theta, \phi_n, W_n}(X_s), f_{\theta, \phi_n, W_n}(X_{t_i}^l))] \\ + \max_i d(f_{\theta, \phi_n, W_n}(X_{t_{i-1}}), f_{\theta, \phi_n, W_n}(X_{t_i}))] \end{aligned} \quad (19)$$

其中, η_{out} 表示外循环的学习率。通过公式 (18) 和 (19) 给出的内外循环优化目标, 可以获得适应不断演变复合域的元表征。当在元测试阶段适应新的目标轨迹 T_{qry} , E_{θ} 的表征是固定不变的, 通过优化 g_{ϕ} 和 c_W 促进模型适应不断变化的目标域数据。为此, 还需要为原始的适配器 g_{ϕ} 配备对应的元适配器 g'_{ϕ} , 以避免模型遗忘从前一个目标域数据上学习的跨域元知识。如果原始适配器可以很好地迁移模型到之前的目标数据, 中间层特征应当包含该目标数据特有知识。因此, 在适应新的目标数据时, 通过匹配中间层特征有助于解决对历史学习跨域知识的遗忘问题。然而, 不同的中间特征在适应新目标数据时发挥的作用有区别, 一个更合理的方案是为不同特征分配不同的权重。

令适配器的第 l 层特征表示为 g_{ϕ_l} , 先前目标域数据的第 l 层特征表示为 $g_{\phi_l}^{pre}$, 加权特征匹配损失可以表示为 $\sum_{j=1}^J \omega_j \|g_{\phi_l}(x_j) - g_{\phi_l}^{pre}(x_j)\|$, 其中 J 为批量样本个数, ω_j 表示样本 x_j 的权重。为了更好地保留跨域知识, 搭建了元适配器 $\omega_j = g'_{\phi_l}(x_j)$, 整体损失是每层加权的特征匹配损失之和, 可以表示为:

$$L_m(\phi', \phi, \phi^{pre}) = \sum_l \sum_j g'_{\phi_l}(x_j) \|g_{\phi_l}(x_j) - g_{\phi_l}^{pre}(x_j)\|. \quad (20)$$

元适配器的加权特征匹配如图 3.8 所示, 在内循环, 通过生成的权重对 g_{ϕ} 和 c_W 进行优化。具体而言, 将特征匹配损失 L_m 加入到公式(18), 并且固定元适配器 g'_{ϕ} 的参数。在外循环中, 需要更新元适配器, 以使加权的特征可以更多地保留先前目标域中特有的知识。在内循环, 训练适配器 g_{ϕ} 和视觉感知预测模块 c_W , 使模型可以适应到连续的目标数据。在外循环, 同时优化编码器 E_{θ} 和元适配器 g'_{ϕ} , 以便促进模型适应不断变化的目标域, 并且不会遗忘从之前目标数据学习到的知识。当元测试阶段遇到新的目标轨迹时, E_{θ} 和 g'_{ϕ} 是固定的, 仅需更新 g_{ϕ} 和 c_W 。通过本部分搭建的复合域适应框架和元表征动态更新方法, 可以使模型在适应不断变化的新域时, 尽可能多地保留历史学习的跨域元知识, 有助于进一步提高对连续域和离散域的广义泛化能力。

3.3 目前研究进展

申请人已经在自动驾驶的跨域视觉感知技术方面开展了以下科研工作:

(1) 构建了一个国内特有的跨环境自动驾驶数据集。自动驾驶系统在实际应用中需要应对复杂的环境变化, 因此, 构建一个大规模的数据集以涵盖不同环境变化和跨域类型, 是进行跨域视觉感知技术研究的基础条件。现有的自动驾驶数据集存在一些不足之处, 例如数据主要采集自国外的驾驶场景、环境变化的多样性不足、天气条件种类少且未涵盖不

同类型的跨域数据、视觉感知任务的标注数量较小等。为了使跨域视觉感知技术更适用于国内自动驾驶场景，申请人在博士后面上基金的资助下，初步构建了国内特有的跨环境视觉感知数据集。与现有数据集的对比结果如表 3.1 所示，可以看出，申请人前期构建的数据集在不同方面进行了改进，尤其在跨域数据分布方面，构建的数据集涵盖了包含离散和连续变化的跨域数据，可以更好地反映自动驾驶跨域数据的独特分布特点。为了更好地实施本项目，需要进一步扩展数据集的规模，扩展视觉感知的多任务标注，并投入更大的力度不断完善前期搭建的国内自动驾驶场景跨域数据集。

表 3.1 自动驾驶场景数据集对比

数据集	来源	天气条件种类	跨域数据分布	任务（标注数据量）
Cityscapes	国外	晴天	无	语义分割（25k）
NuScenes	国外	晴天、雨天及夜间	离散变化域	目标检测（40k） 目标跟踪（40k）
BDD100K	国外	晴天、雨天、雾天、雪天、阴天及夜间	离散变化域	语义分割（10k） 目标检测（100k） 目标跟踪（318k）
本项目前期搭建的数据集	国内	晴天、雨天、雾天、雪天、阴天及夜间	离散+连续变化域	语义分割（40k） 光流估计（40k） 目标检测（120k）

（2）验证模型跨域视觉感知的推理能力。如图 3.8 所示，以语义分割和目标检测为例，申请人在前期搭建的国内自动驾驶场景跨域数据集上，使用 mAP 和 mIoU 作为评估指标验证模型在跨域检测和分割方面的性能，左右两个子图分别对应目标检测结果和分割结果，不同颜色线段表示利用不同环境源域数据训练的模型，线段上的点对应的横坐标表示目标域数据采集的环境。通过预实验结果可以得出结论：一方面，模型在跨域数据上的视觉感知精度下降，另一方面，使用整体数据作为源域训练模型无法有效提高模型在所有目标域上的感知精度。为了提高视觉感知技术的广义泛化能力，保障其在自动驾驶领域的安全性和可靠性，降低不同类型域差异显得至关重要，这也是本项目的研究目标。

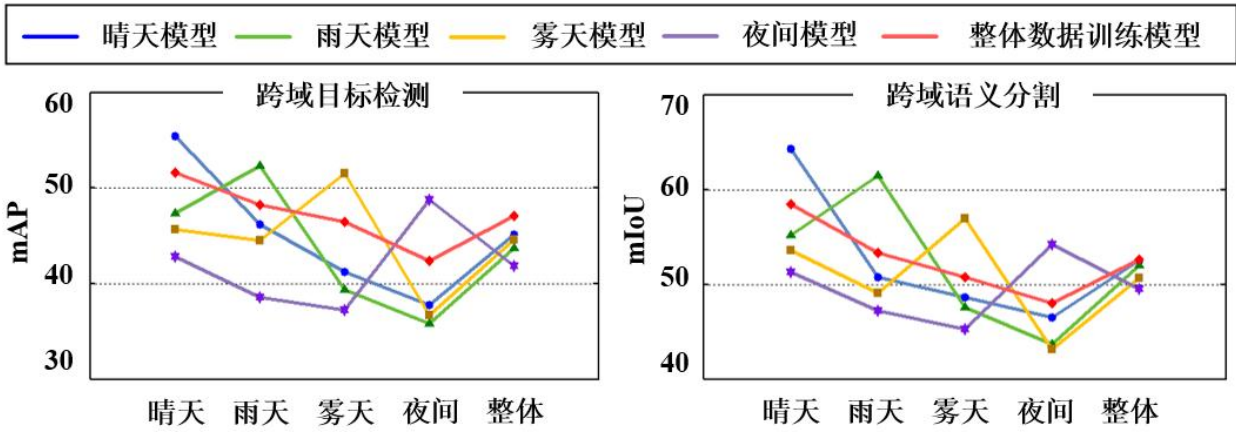


图 3.8 跨域场景的（左图）目标检测和（右图）语义分割结果

3.4 研究计划

2023.07-2023.12: 充分调研和整理相关研究资料, 建立一个专属于国内驾驶场景的复杂跨域数据集, 持续提高数据量并逐步完善数据标注。

2024.01-2024.05: 深入研究跨域视觉感知技术存在的问题, 构建跨连续域的视觉感知框架, 从高维空间中挖掘数据分布特点, 利用元学习驱动连续域分布轨迹聚合。本阶段预计发表 1-2 篇高质量学术论文, 申请 1 项国家发明专利。

2024.05-2024.12: 研究跨域数据的风格迁移方法, 构造基于多配对的元学习离散域适应网络, 降低离散域间迁移视觉感知模型的难度。本阶段预计发表 1-2 篇高质量学术论文, 参加 1-2 次国内外学术会议, 同时申请 1-2 项国家发明专利。

2025.01-2025.07: 搭建基于元表征动态更新的复合域适应框架, 旨在从复杂环境的不同跨域数据中获取高精度的视觉感知结果。本阶段预计发表 2-3 篇高质量学术论文, 参加 1-2 次国内外学术会议, 申请 1-2 项国家发明专利。

3.5 研究基础与工作条件

3.5.1 研究基础

本人长期从事人工智能领域的研究工作, 在计算机视觉技术方面具有较扎实的研究基础, 获得2022年度江苏省人工智能学会优秀博士学位论文奖。丰富的项目经验是完成本项目的关键因素之一, 近年来, 本人作为负责人和核心成员参加了多个科研项目, 其中包括:

(1) 中国博士后科学基金面上项目, 基于鲁棒视觉特征的复杂场景密集预测方法研究, 2022/08-2024-08, 在研, 主持;

(2) 江苏省社会安全图像与视频理解重点实验室创新基金, 基于正交矩的视觉感知方法研究, 2023/01-2023/12, 在研, 主持;

(3) 国家自然科学基金面上项目, 基于异质任务耦合的鲁棒视觉行人搜索, 2022/01-2025/12, 在研, 参与;

(4) 并行与分布处理国防科技重点实验室稳定支持项目, 视频中弱小行人目标的快速检测方法, 2019/05-2020/04, 结题, 参与;

(5) CCF-腾讯犀牛鸟科研基金项目, 基于深度学习的视频中人体分析技术, 2018/09-2019/09, 结题, 参与。

截至目前, 申请人累计发表了 25 篇学术论文, 其中包括 16 篇 SCI 国际期刊, 8 篇国际会议, 1 篇国内一级学报。近三年, 本人以第一作者身份发表学术论文 9 篇, 其中, 5 篇收录于中国计算机学会 (CCF) 推荐的 A 类和 T1 类高质量期刊和会议。本项目以数据分析为出发点, 元知识学习为突破点, 元表征学习为扩展点, 旨在构建一套自动驾驶复杂场景下鲁棒的跨域视觉感知体系。申请人在自动驾驶场景解析、视觉感知以及元学习等领域进行了深入研究, 积累了丰富的理论知识和科研经验, 为本项目的顺利实施打下了坚实的基础。近三年, 申请人主要论文如下:

[1] **本人第一作者**. An accurate and lightweight method for human body image super-resolution. IEEE Transactions on Image Processing (TIP), 2021, 30(2): 2888-2897. (CCF A类期刊, 中科院一区期刊)

- [2] **本人第一作者**. Learning to adapt via latent domains for adaptive semantic segmentation. In: Proceedings of Conference on Neural Information Processing Systems (NeurIPS), 2021. (CCF A类会议)
- [3] **本人第一作者**. Hierarchical information passing based noise-tolerant hybrid learning for semi-supervised human parsing. In: Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2021: 2207-2215. (CCF A类会议)
- [4] **本人第一作者**. Hybrid resolution network using edge guided region mutual information loss for human parsing. In: Proceedings of ACM International Conference on Multimedia (ACM MM), 2020: 1670-1678. (CCF A类会议)
- [5] **本人第一作者**. 基于级联密集网络的轮廓波变换域图像复原. 软件学报, 2020, 31(12): 3968-3980. (CCF T1 类国内高质量科技期刊)
- [6] **本人第一作者**. Single image super-resolution via hybrid resolution NSST prediction. Computer Vision and Image Understanding, 2021, 207(6): 103202. (CCF B类期刊)
- [7] **本人第一作者**. Color image watermark decoder by modeling quaternion polar harmonic transform with BKF distribution. Signal Processing: Image Communication, 2020, 88(10): 115946. (CCF C类期刊)
- [8] **本人第一作者**. Accurate quaternion radial harmonic Fourier moments for color image reconstruction and object recognition. Pattern Analysis and Applications, 2020, 23(4): 1551-1567. (CCF C类期刊)
- [9] **本人第一作者**. Improving image retrieval by integrating shape and texture features. Multimedia Tools and Applications, 2019, 78(2): 2526-2550. (CCF C类期刊)
- [10] **本人第三作者**. Improving image retrieval by integrating shape and texture features. In: Proceedings of ACM International Conference on Multimedia (ACM MM), 2020: 1670-1678. (CCF A类会议)

3. 5. 2 前期研究基础与本项目之间的关系

在申请人的博士后研究期间,得到了中国博士后科学基金面上项目和江苏省社会安全图像和视频理解重点实验室创新基金的支持,开展了以下科研工作:

(1) 为了更好地联合优化不同的视觉感知任务,提出了一种基于自适应权重分配的跨域行人搜索方法,提高了不同场景中行人检测和身份识别的精度;

(2) 为了充分利用不同层次的域相关知识,提出了一种基于多层次知识蒸馏的跨域目标检测方法,提高了目标检测器适应不同场景的能力;

(3) 为了降低域间知识迁移的难度,提出了一种结合风格转换和双层优化的跨域语义分割方法,提高了模型在复杂场景中的语义分割精度;

(4) 发挥稳定正交矩和深度学习的协同优势,提出了一种基于正交矩空间的跨域全景解析方法,提高了解析模型对复杂场景中数据退化的鲁棒性。

在前期的研究工作中,申请人已经围绕复杂场景的视觉感知技术展开了研究,并将相关成果投稿到了 IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, Pattern Recognition 和 Proceedings of the IEEE International Conference on Computer Vision。在本项目中,申请人将进一步关注自动驾驶场景跨域视觉感知技术所面临的挑战。基于前期的研究工作,本项目旨在构建一种高精度的跨域视觉感知系统,提高视觉感知技术在自动驾驶领域的应用价值。本项目是申请人承担的中国博士后科学基金面上项目研究工作的延深与拓展,具有重要的研究意义和应用价值。申请人希望通过本项目的支持,取得更深入和完善的研究成果。

3.5.3 工作条件

本项目依托的****大学智能技术研究中心拥有一支学历层次高、科研能力强的研究队伍，现有人员 20 人，其中教授/博导 3 人，副教授/硕导 11 人，讲师 3 人，在站博士后 3 人，现有在读博士生、硕士生 60 多名。中心先后承担了二十余项国家和省部级科研项目，其中包括国家自然科学基金重大/重点项目子课题 2 项、国家自然科学基金面上/青年/专项等项目 12 项、国家 973 预研项目 1 项，以及企业委托的大型横向项目多项。在国内外重要期刊和国际会议发表论文 300 多篇，其中 SCI 和 EI 收录近 200 篇。部分研究成果在本领域处于国际先进、国内领先的地位。中心成员与其它单位合作取得的研究成果曾获国家科技进步奖二等奖、教育部科技进步奖一等奖、北京市科技进步奖二等奖、辽宁省科技进步二等奖及自然科学学术成果二等奖等。申请人所在的****大学是国家“双一流”重点建设高校，是交通运输部直属的重点高校。多年来，学校高度重视国家级基金项目的申报和实施，在人力、物力和时间上给予了充分的支持，装备各种网络设备、软硬件开发平台等一系列先进的仪器设备和仿真软件。此外，学校拥有齐全的中英文期刊数据库，完全满足本项目研究过程中文献调研的需求。以上均为本项目的顺利开展提供了完备的条件。